



1600

RAW SEQUENCE LISTING

DATE: 02/12/2004

PATENT APPLICATION: US/09/716,964B

TIME: 11:05:12

Input Set : A:\R10301.app

Output Set: N:\CRF4\02122004\I716964B.raw

3 <110> APPLICANT: O'Donnell, Michael E.
 4 Yuzhakov, Alexander
 5 Yurieva, Olga
 6 Jeruzalmi, David
 7 Bruck, Irina
 8 Kuriyan, John
 10 <120> TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT
 11 FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND
 12 USE THEREOF
 14 <130> FILE REFERENCE: 22221/1030
 16 <140> CURRENT APPLICATION NUMBER: 09/716,964B
 17 <141> CURRENT FILING DATE: 2000-11-21
 19 <150> PRIOR APPLICATION NUMBER: 60/143,202
 20 <151> PRIOR FILING DATE: 1997-04-08
 22 <150> PRIOR APPLICATION NUMBER: 08/823,407
 23 <151> PRIOR FILING DATE: 1997-04-08
 25 <150> PRIOR APPLICATION NUMBER: 09/057,416
 26 <151> PRIOR FILING DATE: 1998-04-08
 28 <160> NUMBER OF SEQ ID NOS: 212
 30 <170> SOFTWARE: PatentIn Ver. 2.1
 32 <210> SEQ ID NO: 1
 33 <211> LENGTH: 2007
 34 <212> TYPE: DNA
 35 <213> ORGANISM: Thermus thermophilus
 37 <400> SEQUENCE: 1
 38 tccgggggtg ggggttcccag gtagaccccg gccctcccg tgagccctt tacccaggcc 60
 39 gccacctcct ccaggggggc caaggcgtgc aaggagagga acgtccgcac cacgccctat 120
 40 actagccttg tgagcgccct ctaccgccc ttccgcccc tcaccttcca ggaggtggtg 180
 41 gggcaggagc acgtgaagga gccctcctc aaggccatcc gggaggggag gctcgcccag 240
 42 gcctacctct tctccggggc caggggcgtg ggcaagacca ccacggcgag gctcctcgcc 300
 43 atggcggttg ggtgccagg ggaagacccc ccttgcgggg tctgccccca ctgccaggcg 360
 44 gtgcagaggg gcgcccaccc ggacgtggtg gacattgacg ccgccagcaa caactccgtg 420
 45 gaggacgtgc gggagctgag ggaaaggatc cacctcgccc ccctctctgc ccccaggaag 480
 46 gtcttcatcc tggacgaggc ccacatgctc tccaaaagcg ccttcaacgc cctcctcaag 540
 47 accctggagg agcccccgcc ccacgtcctc ttctgttctg ccaccaccga gcccgagagg 600
 48 atgcccccca ccacctctc ccgcacccag cacttccgct tccgcccgt cacggaggag 660
 49 gagatcgctt ttaagctccg gcgcacctg gaggcggtg ggcgggaggc ggaggaggag 720
 50 gccctcctcc tcctcgccc cctggcggag ggggccccta gggacgcgga aagcctcctg 780
 51 gaggccttcc tcctcgtgga agggcccctc acccggaagg aggtggagcg cgccctaggc 840
 52 tccccccag ggaccgggtt ggccgagatc gccgcctccc tcgcgagggg gaaaacggcg 900
 53 gaggcccttg gcctcgccc gcgcctctac ggggaagggt acgccccgag gagcctggtc 960
 54 tcgggccttt tggagggtgt ccgggaaggc ctctacgcg ccttcggcct cgcggaacc 1020
 55 ccccttccc ccccgcccca ggccctgatc gccgccatga ccgccctgga cgaggccatg 1080

p.6

ENTERED

RAW SEQUENCE LISTING

DATE: 02/12/2004

PATENT APPLICATION: US/09/716,964B

TIME: 11:05:12

Input Set : A:\R10301.app

Output Set: N:\CRF4\02122004\I716964B.raw

```

56 gagcgccctcg cccgccgctc cgacgcctta agcctggagg tggccctcct ggaggcgga 1140
57 agggccctgg ccgccgaggc cctaccccag cccacgggcg ctccttcccc agaggctggc 1200
58 cccaagccgg aaagcccccc gaccccgga ccccaaggc ccgaggaggc gcccgaacctg 1260
59 cgggagcggt ggcgggcctt cctcgaggcc ctcaggccca ccctacgggc cttcgtgcgg 1320
60 gagggcccgc cgagggtccg ggaaggccag ctctgcctcg ctttccccga ggacaaggcc 1380
61 ttccactacc gcaaggcctc ggaacagaag gtgaggctcc tccccctggc ccaggcccat 1440
62 ttcgggggtg aggaggctcg cctcgctctg gagggagaaa aaaaaagcct gagcccaagg 1500
63 ccccgcccg cccacactcc tgaagcgccc gcaccccg ggcctcccga ggaggaggta 1560
64 gaggcggagg aagcggcgga ggaggcccc gaggaggcct tgaggcggtt ggtccgcctc 1620
65 ctgggggggc ggggtgctct ggtgcggcg cccaggacc gggaggcgcc ggaggaggaa 1680
66 cccctgagcc aagacgagat agggggtact ggtatataat gggggcatga cgcggaccac 1740
67 cgacctcgga caagagaccg tggacaacat cctcaagcgc ctccgccgta ttgagggcca 1800
68 ggtgcggggg gccagaaga tgggtggccga gggccgcccc tgcgacgagg tcctcaccca 1860
69 gatgaccgcc accaagaagg ccatggaggc ggcggccacc ctgatcctcc acgagttcct 1920
70 gaacgtctgc gccgccgagg tctccgagg caaggtgaac cccaagaagc ccgaggagat 1980
71 cgccaccatg ctgaagaact tcatcta 2007
74 <210> SEQ ID NO: 2
75 <211> LENGTH: 529
76 <212> TYPE: PRT
77 <213> ORGANISM: Thermus thermophilus
79 <400> SEQUENCE: 2
80 Met Ser Ala Leu Tyr Arg Arg Phe Arg Pro Leu Thr Phe Gln Glu Val
81 1 5 10 15
83 Val Gly Gln Glu His Val Lys Glu Pro Leu Leu Lys Ala Ile Arg Glu
84 20 25 30
86 Gly Arg Leu Ala Gln Ala Tyr Leu Phe Ser Gly Pro Arg Gly Val Gly
87 35 40 45
89 Lys Thr Thr Thr Ala Arg Leu Leu Ala Met Ala Val Gly Cys Gln Gly
90 50 55 60
92 Glu Asp Pro Pro Cys Gly Val Cys Pro His Cys Gln Ala Val Gln Arg
93 65 70 75 80
95 Gly Ala His Pro Asp Val Val Asp Ile Asp Ala Ala Ser Asn Asn Ser
96 85 90 95
98 Val Glu Asp Val Arg Glu Leu Arg Glu Arg Ile His Leu Ala Pro Leu
99 100 105 110
101 Ser Ala Pro Arg Lys Val Phe Ile Leu Asp Glu Ala His Met Leu Ser
102 115 120 125
104 Lys Ser Ala Phe Asn Ala Leu Leu Lys Thr Leu Glu Glu Pro Pro Pro
105 130 135 140
107 His Val Leu Phe Val Phe Ala Thr Thr Glu Pro Glu Arg Met Pro Pro
108 145 150 155 160
110 Thr Ile Leu Ser Arg Thr Gln His Phe Arg Phe Arg Arg Leu Thr Glu
111 165 170 175
113 Glu Glu Ile Ala Phe Lys Leu Arg Arg Ile Leu Glu Ala Val Gly Arg
114 180 185 190
116 Glu Ala Glu Glu Glu Ala Leu Leu Leu Leu Ala Arg Leu Ala Asp Gly
117 195 200 205
119 Ala Leu Arg Asp Ala Glu Ser Leu Leu Glu Arg Phe Leu Leu Leu Glu
120 210 215 220

```

RAW SEQUENCE LISTING

DATE: 02/12/2004

PATENT APPLICATION: US/09/716,964B

TIME: 11:05:12

Input Set : A:\R10301.app

Output Set: N:\CRF4\02122004\I716964B.raw

```

122 Gly Pro Leu Thr Arg Lys Glu Val Glu Arg Ala Leu Gly Ser Pro Pro
123 225                230                235                240
125 Gly Thr Gly Val Ala Glu Ile Ala Ala Ser Leu Ala Arg Gly Lys Thr
126                245                250                255
128 Ala Glu Ala Leu Gly Leu Ala Arg Arg Leu Tyr Gly Glu Gly Tyr Ala
129                260                265                270
131 Pro Arg Ser Leu Val Ser Gly Leu Leu Glu Val Phe Arg Glu Gly Leu
132                275                280                285
134 Tyr Ala Ala Phe Gly Leu Ala Gly Thr Pro Leu Pro Ala Pro Pro Gln
135                290                295                300
137 Ala Leu Ile Ala Ala Met Thr Ala Leu Asp Glu Ala Met Glu Arg Leu
138 305                310                315                320
140 Ala Arg Arg Ser Asp Ala Leu Ser Leu Glu Val Ala Leu Leu Glu Ala
141                325                330                335
143 Gly Arg Ala Leu Ala Ala Glu Ala Leu Pro Gln Pro Thr Gly Ala Pro
144                340                345                350
146 Ser Pro Glu Val Gly Pro Lys Pro Glu Ser Pro Pro Thr Pro Glu Pro
147                355                360                365
149 Pro Arg Pro Glu Glu Ala Pro Asp Leu Arg Glu Arg Trp Arg Ala Phe
150                370                375                380
152 Leu Glu Ala Leu Arg Pro Thr Leu Arg Ala Phe Val Arg Glu Ala Arg
153 385                390                395                400
155 Pro Glu Val Arg Glu Gly Gln Leu Cys Leu Ala Phe Pro Glu Asp Lys
156                405                410                415
158 Ala Phe His Tyr Arg Lys Ala Ser Glu Gln Lys Val Arg Leu Leu Pro
159                420                425                430
161 Leu Ala Gln Ala His Phe Gly Val Glu Glu Val Val Leu Val Leu Glu
162                435                440                445
164 Gly Glu Lys Lys Ser Leu Ser Pro Arg Pro Arg Pro Ala Pro Pro Pro
165                450                455                460
167 Glu Ala Pro Ala Pro Pro Gly Pro Pro Glu Glu Glu Val Glu Ala Glu
168 465                470                475                480
170 Glu Ala Ala Glu Glu Ala Pro Glu Glu Ala Leu Arg Arg Val Val Arg
171                485                490                495
173 Leu Leu Gly Gly Arg Val Leu Trp Val Arg Arg Pro Arg Thr Arg Glu
174                500                505                510
176 Ala Pro Glu Glu Glu Pro Leu Ser Gln Asp Glu Ile Gly Gly Thr Gly
177                515                520                525

```

179 Ile

183 <210> SEQ ID NO: 3

184 <211> LENGTH: 1590

185 <212> TYPE: DNA

186 <213> ORGANISM: Thermus thermophilus

188 <400> SEQUENCE: 3

```

189 gtgagcgccc tctaccgccc cttccgcccc ctcaccttcc aggaggtggt ggggcaggag 60
190 cagtgaaagg agccctcctt caaggccatc cgaggaggga ggctcgccca ggcctacctc 120
191 ttctccgggc ccaggggcgt gggcaagacc accacggcga ggctcctcgc catggcggtg 180
192 ggggtgccagg ggggaagacc cccttgcggg gtctgcccc actgccaggc ggtgcagagg 240
193 ggcgcccacc cggacgtggt ggacattgac gccgccagca acaactccgt ggaggacgtg 300

```

RAW SEQUENCE LISTING

DATE: 02/12/2004

PATENT APPLICATION: US/09/716,964B

TIME: 11:05:12

Input Set : A:\R10301.app

Output Set: N:\CRF4\02122004\I716964B.raw

```

194 cgggagctga gggaaaggat ccacctcgcc cccctctctg cccccaggaa ggtcttcac 360
195 ctggacgagg cccacatgct ctccaaaagc gccttcaacg cctcctcaa gacctggag 420
196 gagccccgcg cccacgtcct ctctgtcttc gccaccaccg agcccgagag gatgcccccc 480
197 accatcctct cccgcaccca gcacttcgcg ttccgccgcc tcacggagga ggagatcgcc 540
198 tttaaagctcc ggcgcacatc ggaggccgtg gggcgggagg cgaggaggga ggccctcctc 600
199 ctctcgcgcc gcctggcgga cggggccctt agggacgcgg aaagcctcct ggagcgcttc 660
200 ctctccttg aaggccccct cacccggaag gaggtggagc gcgccttagg ctccccccca 720
201 gggaccgggg tggccgagat cgccgcctcc ctcgcgaggg ggaacacggc ggaggccctg 780
202 ggccctcgcc ggcgcctcta cggggaaggg tacgccccga ggagcctggt ctcgggcctt 840
203 ttggagggtg tccgggaagg cctctacgcc gccttcggcc tcgcgggaac ccccttccc 900
204 gccccgcccc aggcctgat cgccgccatg accgccttg acgaggccat ggagcgctc 960
205 gcccgcgctt ccgacgcctt aagcctggag gtggccctcc tggaggcggg aagggccctg 1020
206 gccgcccagg cctaccccc gcccacgggc gctccttccc cagaggtcgg cccaagccg 1080
207 gaaagccccc cgacccccga acccccaagg cccgaggagg cgcgcgacct gcgggagcgg 1140
208 tggcgggcct tctcagagg cctcaggccc accctacggg ccttcgtgcg ggaggccgc 1200
209 ccggagggtc gggaaggcca gctctgcctc gctttccccg aggacaaggc cttccactac 1260
210 cgcaaggcct cggaacagaa ggtgaggctc ctccccctgg ccagggccca tttcggggtg 1320
211 gaggagggtc tctcgtcct ggaggagaa aaaaaagcc tgagcccaag gccccgccc 1380
212 gccccacctc ctgaagcgcc cgcacccccg ggccctccc aggaggagg agaggcgag 1440
213 gaagcggcgg aggaggcccc ggaggaggcc ttgaggcggg tggtcgcct cctggggggg 1500
214 cgggtgctct gggcgcgcg gcccaggacc cgggaggcgc cgaggaggga acccctgagc 1560
215 caagacgaga taggggtac tggatatataa 1590

```

218 <210> SEQ ID NO: 4

219 <211> LENGTH: 464

220 <212> TYPE: PRT

221 <213> ORGANISM: Thermus thermophilus

223 <400> SEQUENCE: 4

```

224 Met Ser Ala Leu Tyr Arg Arg Phe Arg Pro Leu Thr Phe Gln Glu Val
225 1 5 10 15
227 Val Gly Gln Glu His Val Lys Glu Pro Leu Leu Lys Ala Ile Arg Glu
228 20 25 30
230 Gly Arg Leu Ala Gln Ala Tyr Leu Phe Ser Gly Pro Arg Gly Val Gly
231 35 40 45
233 Lys Thr Thr Thr Ala Arg Leu Leu Ala Met Ala Val Gly Cys Gln Gly
234 50 55 60
236 Glu Asp Pro Pro Cys Gly Val Cys Pro His Cys Gln Ala Val Gln Arg
237 65 70 75 80
239 Gly Ala His Pro Asp Val Val Asp Ile Asp Ala Ala Ser Asn Asn Ser
240 85 90 95
242 Val Glu Asp Val Arg Glu Leu Arg Glu Arg Ile His Leu Ala Pro Leu
243 100 105 110
245 Ser Ala Pro Arg Lys Val Phe Ile Leu Asp Glu Ala His Met Leu Ser
246 115 120 125
248 Lys Ser Ala Phe Asn Ala Leu Leu Lys Thr Leu Glu Glu Pro Pro Pro
249 130 135 140
251 His Val Leu Phe Val Phe Ala Thr Thr Glu Pro Glu Arg Met Pro Pro
252 145 150 155 160
254 Thr Ile Leu Ser Arg Thr Gln His Phe Arg Phe Arg Arg Leu Thr Glu
255 165 170 175

```

RAW SEQUENCE LISTING

DATE: 02/12/2004

PATENT APPLICATION: US/09/716,964B

TIME: 11:05:12

Input Set : A:\R10301.app

Output Set: N:\CRF4\02122004\I716964B.raw

```

257 Glu Glu Ile Ala Phe Lys Leu Arg Arg Ile Leu Glu Ala Val Gly Arg
258      180      185      190
260 Glu Ala Glu Glu Glu Ala Leu Leu Leu Leu Ala Arg Leu Ala Asp Gly
261      195      200      205
263 Ala Leu Arg Asp Ala Glu Ser Leu Leu Glu Arg Phe Leu Leu Leu Glu
264      210      215      220
266 Gly Pro Leu Thr Arg Lys Glu Val Glu Arg Ala Leu Gly Ser Pro Pro
267 225      230      235      240
269 Gly Thr Gly Val Ala Glu Ile Ala Ala Ser Leu Ala Arg Gly Lys Thr
270      245      250      255
272 Ala Glu Ala Leu Gly Leu Ala Arg Arg Leu Tyr Gly Glu Gly Tyr Ala
273      260      265      270
275 Pro Arg Ser Leu Val Ser Gly Leu Leu Glu Val Phe Arg Glu Gly Leu
276      275      280      285
278 Tyr Ala Ala Phe Gly Leu Ala Gly Thr Pro Leu Pro Ala Pro Pro Gln
279      290      295      300
281 Ala Leu Ile Ala Ala Met Thr Ala Leu Asp Glu Ala Met Glu Arg Leu
282 305      310      315      320
284 Ala Arg Arg Ser Asp Ala Leu Ser Leu Glu Val Ala Leu Leu Glu Ala
285      325      330      335
287 Gly Arg Ala Leu Ala Ala Glu Ala Leu Pro Gln Pro Thr Gly Ala Pro
288      340      345      350
290 Ser Pro Glu Val Gly Pro Lys Pro Glu Ser Pro Pro Thr Pro Glu Pro
291      355      360      365
293 Pro Arg Pro Glu Glu Ala Pro Asp Leu Arg Glu Arg Trp Arg Ala Phe
294      370      375      380
296 Leu Glu Ala Leu Arg Pro Thr Leu Arg Ala Phe Val Arg Glu Ala Arg
297 385      390      395      400
299 Pro Glu Val Arg Glu Gly Gln Leu Cys Leu Ala Phe Pro Glu Asp Lys
300      405      410      415
302 Ala Phe His Tyr Arg Lys Ala Ser Glu Gln Lys Val Arg Leu Leu Pro
303      420      425      430
305 Leu Ala Gln Ala His Phe Gly Val Glu Glu Val Val Leu Val Leu Glu
306      435      440      445
308 Gly Glu Lys Lys Lys Pro Glu Pro Lys Ala Pro Pro Gly Pro Thr Ser
309      450      455      460
315 <210> SEQ ID NO: 5
316 <211> LENGTH: 454
317 <212> TYPE: PRT
318 <213> ORGANISM: Thermus thermophilus
320 <400> SEQUENCE: 5
321 Met Ser Ala Leu Tyr Arg Arg Phe Arg Pro Leu Thr Phe Gln Glu Val
322 1      5      10      15
324 Val Gly Gln Glu His Val Lys Glu Pro Leu Leu Lys Ala Ile Arg Glu
325      20      25      30
327 Gly Arg Leu Ala Gln Ala Tyr Leu Phe Ser Gly Pro Arg Gly Val Gly
328      35      40      45
330 Lys Thr Thr Thr Ala Arg Leu Leu Ala Met Ala Val Gly Cys Gln Gly
331      50      55      60

```

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 02/12/2004
PATENT APPLICATION: US/09/716,964B TIME: 11:05:13

Input Set : A:\R10301.app
Output Set: N:\CRF4\02122004\I716964B.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:17; Xaa Pos. 2,3,5
Seq#:29; N Pos. 6,12,21
Seq#:30; N Pos. 7,10,19,22
Seq#:42; N Pos. 7,8,13,14
Seq#:43; N Pos. 8,9,17,18
Seq#:66; Xaa Pos. 3,5
Seq#:67; Xaa Pos. 4,7
Seq#:68; Xaa Pos. 3,5
Seq#:89; Xaa Pos. 79
Seq#:91; Xaa Pos. 47,57

VERIFICATION SUMMARY

DATE: 02/12/2004

PATENT APPLICATION: US/09/716,964B

TIME: 11:05:13

Input Set : A:\R10301.app

Output Set: N:\CRF4\02122004\I716964B.raw

L:567 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:17 after pos.:0
L:1092 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:29 after pos.:0
L:1124 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:30 after pos.:0
L:1288 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:42 after pos.:0
L:1320 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:43 after pos.:0
L:1594 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:66 after pos.:0
L:1617 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:67 after pos.:0
L:1640 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:68 after pos.:0
L:2282 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:89 after pos.:64
L:2375 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:91 after pos.:32
M:341 Repeated in SeqNo=91